Testing Shape Restrictions with Continuous Treatment: A Transformation Model Approach*

Arkadiusz Szydłowski†

University of Leicester

[PLEASE SEE `https://arekszydlowski.github.io` FOR THE LATEST VERSION]

November 7, 2022

Abstract

We propose tests for the convexity/concavity of transformation of the dependent variable in a semiparametric transformation model. The tests can be used to verify monotonicity of the treatment effect, or, in other words, concavity/convexity of the outcome with respect to the treatment, in (quasi-)experimental settings. Our procedure does not require estimation of the transformation or the distribution of the error terms, thus it is easy to implement. The statistic has a form of a U statistic or a localised U statistic and we show that critical values can be obtained by bootstrapping. In our application we test convexity of loan demand in interest rate.

JEL: C12, C21, C14

Keywords: Shape restrictions, Treatment effects, Transformation model, Bootstrap, U statistic

*I would like to thank Nikolas Mittag and Pedro Sant'Anna for useful discussions and comments. This research used the ALICE High Performance Computing Facility at the University of Leicester.

†Department of Economics, Finance and Accounting, University of Leicester, University Road, Leicester LE1 7RH, UK. *E-mail address*: ams102@le.ac.uk

1 Introduction

In this paper we consider testing if a treatment has a diminishing or increasing effect on the outcome. This is often of interest on top of the question if there is an effect or what sign it has. For example, often it is natural to expect that demand will be decreasing in price. However, it is less clear if increasing the price will have a larger effect at lower or higher price levels. Our test can address this question without fully estimating the demand relationship.

Let X denote the vector of treatments and control variables. As estimating the effect of X on the outcome Y nonparametrically would suffer from a curse of dimensionality with non-trivial number of treatments and controls, we impose a single-index structure and assume that the nonlinearity of the treatment effect comes from a nonlinear transformation of Y. In other words, consider a transformation model of the form:

$$T(Y) = X'\beta + \varepsilon \tag{1}$$

where Y is a scalar dependent variable, X is a vector of q nondegenerate explanatory variables, β is a vector of coefficients belonging to a compact set $\Theta_\beta \subset \mathbb{R}^q$, $T(\cdot)$,increasing function with $T(0) = 0$ (normalization) and ε is an unobserved error term with distribution F that is independent of X. The benefit of using the transformation model compared to a standard single-index model (i.e. $Y = T(X'\beta) + \varepsilon$), besides the fact that it facilitates our testing approach, is that the transformation model allows the treatment effect of X_k to depend on the values of both observed and unobserved characteristics (note that $Y = T^{-1}(X'\beta + \varepsilon)$) so can be seen as a simple way of introducing heterogenous treatment effects.

The main objective of this article is to provide a practically appealing test to verify if the transformation function $T(\cdot)$ is concave/linear/convex. The examples below illustrate the importance of testing curvature of the transformation.

Example 1. (Experiments with continuous treatment) *Using normalization $E(\varepsilon) = 0$ and $Q_\alpha(\varepsilon) = 0$ where Q_α denotes α quantile, we have, respectively:*

$$\frac{\partial^2 E(Y|X)}{\partial X_k^2} = -\beta_k^2 E\left[\frac{T''(T^{-1}(X'\beta + \varepsilon))}{T'(T^{-1}(X'\beta + \varepsilon))^3}\middle| X\right] \qquad \textit{(mean regression)}$$

and

$$\frac{\partial^2 Q_\alpha(Y|X)}{\partial X_k^2} = -\beta_k^2 \frac{T''(T^{-1}(X'\beta))}{T'(T^{-1}(X'\beta))^3} \qquad \textit{(quantile regression)}$$

As $T'(\cdot) > 0$ the curvature of the mean/quantile effect depends on $T''(\cdot)$. Thus, the sign of the second derivative $T''(\cdot)$ determines if the effect of the treatment X on the α quantile of Y is concave or convex in X. For example, if a company randomises marketing spending in different markets, testing for concavity would answer the question if marketing spending has diminishing returns (e.g. on mean revenue). Test of curvature has also application in experimental studies of demand elasticities (e.g. Jessoe & Rapson (2014), Hainmueller et al. (2015), Karlan & Zinman (2018)) where it can be used to verify if demand is concave, linear or convex.

Finally, when the transformation is linear, the treatment effect does not depend on the observed (X) and unobserved (ε) heterogeneity. Thus, the test of linearity of T can be seen as a test of treatment effect heterogeneity.

Example 2. (Duration models: testing hazard monotonicity) *Let $\lambda(\cdot)$ and $\Lambda(\cdot)$ denote baseline hazard and integrated baseline hazard respectively. In a duration model $T(Y) = \log \Lambda(Y)$ and*

$$T''(Y) = \frac{\lambda'(Y)}{\Lambda(Y)} - \left(\frac{\lambda(Y)}{\Lambda(Y)}\right)^2.$$

Hence, rejecting concavity of $T(\cdot)$ (i.e. $T''(\cdot) < 0$) implies that the baseline hazard is non-decreasing ($\lambda'(\cdot) \geq 0$). One can, thus, use the test of concavity of the transformation as a test for monotonicity of the baseline hazard, or in other words, as a test for positive duration dependence. In the economic context, one may be interested in detecting non-monotonicity of unemployment exit rate due to unemployment benefit exhaustion effects (see Card et al. (2007) for discussion).

Beyond these examples our procedure can also be used for specification search, i.e. determining if one should use concave or convex transformation, and to test the curvature in wage regressions, e.g. if the effect of education or experience is concave, or the curvature of the marginal utility (profit) function in hedonic models (see Ekeland et al. (2004)), if the assumption of selection on observables is met.

Our test statistic simply compares triples of Y's corresponding to equally spaced index values $X'\beta$, thus it does not require estimation of the transformation function T or the distribution of ε. We only require an estimator of β and symmetry of the distribution of the error terms. As the null hypothesis in our test assumes linearity, estimating β under the null can be done simply by OLS (alternatively, one can use the maximum rank correlation estimator, Han (1987), or semiparametric least squares, Ichimura (1993)).

We propose both a global test that has power to detect globally convex or concave functions and leads to asymptotic standard normal critical values, as well as a more general test that detects local deviations from linearity, i.e. has power against alternatives that are both convex and concave on the domain of T. The local test does not have a pivotal asymptotic distribution but we show that the critical values can be obtained by bootstrap.

Our statistic resembles the approach in Abrevaya & Jiang (2005) who test curvature in a non-parametric regression model. However, unlike their approach our test does not suffer from the curse of dimensionality due to the single index structure of the regression part. Also the details of the derivation of the asymptotic distribution are different due to presence of estimated β in our statistic and somewhat distinct approach to obtaining power against general alternatives. These traits are shared by Abrevaya et al. (2010) and our derivations follow similar lines to their article, though a distinguishing feature is that we formally show validity of bootstrap for obtaining critical values.

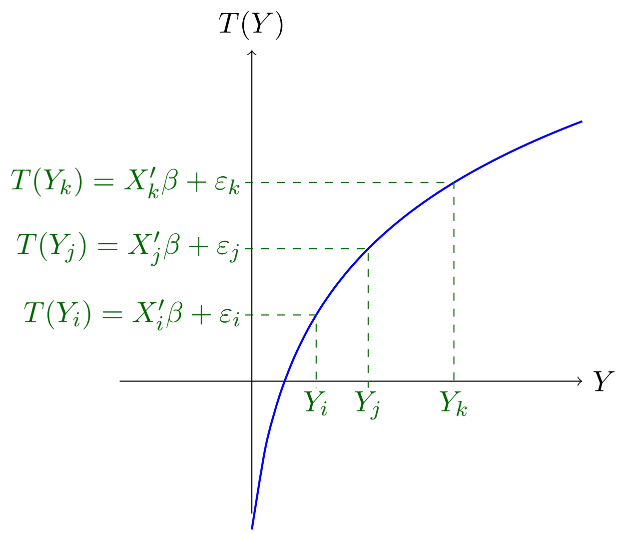
Related literature includes tests for the sign of the treatment effect, see e.g. Kline (2016). Testing curvature of the transformation can be also seen as a generalisation of specification testing in Neumeyer et al. (2016) and Szydłowski (2020). The idea of using curvature of integrated hazard to test monotonicity of baseline hazard has been utilised by Hall & Keilegom (2005). Testing shape restrictions in a nonparametric regression model has been considered by Ghosal et al. (2000), Gutknecht (2016), Chetverikov (2019) and Komarova & Hidalgo (2020), among others.

The article is organized as follows. Section 2 discusses the idea behind the testing procedure informally and formal results are postponed till Sections 3-4. Sections 5-6 contain Monte Carlo results and our application to loan demand. Proofs, besides the proof of the main proposition, are located in the Appendix.

2 Main idea

Figure 2 portrays the intuition behind our test. The transformation plotted in the figure is concave and we display three (ordered) data points in the figure (Y_i, Y_j, Y_k) for which the transformation functions is equally spaced, i.e. $T(Y_k) - T(Y_j) = T(Y_j) - T(Y_i)$.

Figure 1: Testing concavity



Concavity of $T(\cdot)$ implies that $Y_k - Y_j > Y_j - Y_i$. Note that $T(Y_k) - T(Y_j) = T(Y_j) - T(Y_i)$ is equivalent to $(X_k - X_j)'\beta + \varepsilon_k - \varepsilon_j = (X_j - X_i)'\beta + \varepsilon_j - \varepsilon_i$. Hence, "on average" equally spaced $T(Y)$'s mean equally spaced index values $X'\beta$. Therefore, we can detect deviations from concavity by considering the following criterion:

$$-\frac{1}{n(n-1)(n-2)} \sum_{i \neq j \neq k} \mathbb{1}\{Y_i < Y_j < Y_k\} \mathbb{1}\{Y_k - Y_j < Y_j - Y_i\} \mathbb{1}\{X_{kj}'\beta = X_{ji}'\beta\}$$

where $X_{ji} \equiv X_j - X_i$.

In order to make this criterion operational with continuous distribution of $X'\beta$ (which is required for identification) we need to replace the last indicator function with a smooth kernel $K_h(\cdot) = h^{-1}K(\cdot/h)$ and β with its estimator $\hat{\beta}$:

$$-\frac{1}{n(n-1)(n-2)} \sum_{i \neq j \neq k} \mathbb{1}\{Y_i < Y_j < Y_k\} \mathbb{1}\{Y_k - Y_j < Y_j - Y_i\} K_h\left(X_{kj}'\hat{\beta} - X_{ji}'\hat{\beta}\right)$$

Deviations from convexity can be detected in a similar fashion. Finally, we can combine both criterion functions to detect deviations from linearity:

$$U_n = \frac{1}{n(n-1)(n-2)} \sum_{i \neq j \neq k} \mathbb{1}\{Y_i < Y_j < Y_k\} sgn(Y_k - 2Y_j + Y_i) K_h \left(X'_{kj}\hat{\beta} - X'_{ji}\hat{\beta} \right) \tag{2}$$

Here, very negative values of the objective function signify convexity and large positive values mark concavity. Outside testing, one may use the above measure itself to characterise an "average" curvature of function T, e.g. large positive values suggest that the function is predominantly concave.

Note that our test requires only one-dimensional kernel smoothing, thus it does not suffer from the curse of dimensionality. Also, unlike the specification test in Szydłowski (2020) it does not require estimation of the transformation function, a computationally intense task itself.

3 Formal definition and asymptotic theory

3.1 Global test

As the probability limit of the criterion functions introduced in the previous section depends on the extent of concavity, but these functions are centred at known values under linearity, we form our procedure as the test of:

$$H_0 : \begin{cases} T(\cdot) \text{ is concave} \\ T(\cdot) \text{ is linear} \\ T(\cdot) \text{ is convex} \end{cases} \qquad vs \qquad H_A : \begin{cases} T(\cdot) \text{ is non-concave} \\ T(\cdot) \text{ is non-linear} \\ T(\cdot) \text{ is non-convex} \end{cases}$$

depending on the question of interest. We will use the test statistic

$$S_n = \sqrt{n} U_n$$

and reject the null hypothesis at level α if $S_n < c_\alpha, |S_n| > c_{1-\alpha/2}$ or $S_n > c_{1-\alpha}$ respectively, depending on H_0, where c_α denotes an α quantile from an appropriate asymptotic distribution. Proposition 1 justifies our testing strategy.

Proposition 1. *Assume that the distribution of ε is symmetric and that $\{(X_i, Y_i)\}_{i=1}^n$ are i.i.d. Then, as $n \to \infty$, $U_n \to^p \theta$ where: (i) $\theta \geq 0$ if $T(\cdot)$ is globally concave, (ii) $\theta = 0$ if $T(\cdot)$ is globally linear, (iii) $\theta \leq 0$ if $T(\cdot)$ is globally convex.*

Proof. By standard arguments:

$$\begin{aligned} U_n \to^p \quad & E[\mathbb{1}\{Y_i < Y_j < Y_k\} sgn(Y_k - 2Y_j + Y_i) \mathbb{1}\{X'_{kj}\beta = X'_{ji}\beta\}] \\ & = E[sgn(Y_k - 2Y_j + Y_i) | Y_i < Y_j < Y_k, X'_{kj}\beta = X'_{ji}\beta] P(Y_i < Y_j < Y_k, X'_{kj}\beta = X'_{ji}\beta) \\ & \equiv \tilde{\theta} P(Y_i < Y_j < Y_k, X'_{kj}\beta = X'_{ji}\beta) \end{aligned}$$

Note that:

$$\begin{aligned} \tilde{\theta} &= P(Y_k - 2Y_j + Y_i > 0 | Y_i < Y_j < Y_k, X'_{kj}\beta = X'_{ji}\beta) - P(Y_k - 2Y_j + Y_i < 0 | Y_i < Y_j < Y_k, X'_{kj}\beta = X'_{ji}\beta) \\ &\equiv (1) - (2) \end{aligned}$$

and the sign of the probability limit of U_n is determined by the sign of $\tilde{\theta}$. For simplicity let Ξ denote the conditioning event in the probabilities above.

(i) We will show that (1) $\geq$ (2) under concavity of T (i.e. convexity of T^{-1}). We have:

$$\begin{aligned} P(Y_k - 2Y_j + Y_i > 0 | \Xi) &= \int P(Y_k - 2Y_j + Y_i > 0 | Y_i < Y_j < Y_k, X'_{kj}\beta = X'_{ji}\beta = \xi) dF(\xi) \\ &\equiv \int P(Y_k - 2Y_j + Y_i > 0 | \Xi(\xi)) dF(\xi) \end{aligned}$$

Let $\varepsilon_{ji} \equiv \varepsilon_j - \varepsilon_i$. Denote $a = X'_j\beta + \varepsilon_j$, $b = X'_i\beta + \varepsilon_i$ and $\Delta = \xi + \varepsilon_{kj}$. Note that the conditioning event $Y_i < Y_j < Y_k$ implies that $\Delta > 0, a - b > 0$ by monotonicity of T. We can rewrite the event $Y_k - 2Y_j + Y_i > 0$ as:

$$T^{-1}(a + \Delta) - T^{-1}(a) > T^{-1}(a) - T^{-1}(b)$$

Conditional on $\Xi(\xi)$ this event is implied by $\varepsilon_{kj} > \varepsilon_{ji}$. To see that observe that the latter event implies $\Delta > a - b$ which under convexity of T^{-1} (i.e. concavity of T) gives the desired result. Thus, we have:

$$P(Y_k - 2Y_j + Y_i > 0|\Xi) \geq P(\varepsilon_{kj} > \varepsilon_{ji}|\Xi).$$

On the other hand, $Y_k - 2Y_j + Y_i < 0 \iff T^{-1}(a+\Delta) - T^{-1}(a) < T^{-1}(a) - T^{-1}(b)$, which under concavity implies $\varepsilon_{kj} < \varepsilon_{ji}$ as we need $\Delta < a - b$ for this event to occur and by definition $\Delta = a - b + \varepsilon_{kj} - \varepsilon_{ji}$. Hence:

$$P(Y_k - 2Y_j + Y_i < 0|\Xi) \leq P(\varepsilon_{kj} < \varepsilon_{ji}|\Xi).$$

which implies $\tilde{\theta} \geq 2P(\varepsilon_{kj} > \varepsilon_{ji}|\Xi) - 1$ and in order to show that $\tilde{\theta} \geq 0$ we need to show that $P(\varepsilon_{kj} < \varepsilon_{ji}|\Xi) = 0.5$.

Let $F(\cdot)$ denote the distribution of ε. Direct calculation gives:

$$\begin{aligned} P(\varepsilon_{kj} < \varepsilon_{ji}|Y_i < Y_j < Y_k, X'_{kj}\beta = X'_{ji}\beta = \xi) =& \frac{P(\varepsilon_{kj} < \varepsilon_{ji}, Y_i < Y_j < Y_k|X'_{kj}\beta = X'_{ji}\beta = \xi)}{P(Y_i < Y_j < Y_k|X'_{kj}\beta = X'_{ji}\beta = \xi)} \\ =& \frac{\int\int_{-\infty}^{\varepsilon_2+\xi}[F(2\varepsilon_2 - \varepsilon_1) - F(\varepsilon_2 - \xi)]dF(\varepsilon_1)dF(\varepsilon_2)}{\int F(\varepsilon_2+\xi)F(-\varepsilon_2+\xi)dF(\varepsilon_2)} \end{aligned}$$

Note that under the symmetry of the distribution of ε:

$$\begin{aligned} \int\int_{\infty}^{\varepsilon_2+\xi}[F(2\varepsilon_2 - \varepsilon_1) - F(\varepsilon_2 - \xi)]dF(\varepsilon_1)dF(\varepsilon_2) =& -\int\int_{-\infty}^{\varepsilon_2+\xi} F(\varepsilon_1 - 2\varepsilon_2)dF(\varepsilon_1)dF(\varepsilon_2) \\ &+ \int F(\varepsilon_2+\xi)F(-\varepsilon_2+\xi)dF(\varepsilon_2) \end{aligned}$$

and $\int F(\varepsilon_2+\xi)F(-\varepsilon_2+\xi)dF(\varepsilon_2) = 2\int_0^\infty F(\varepsilon_2+\xi)F(-\varepsilon_2+\xi)dF(\varepsilon_2)$. Thus, showing that $P(\varepsilon_{kj} < \varepsilon_{ji}|Y_i < Y_j < Y_k, X'_{kj}\beta = X'_{ji}\beta = \xi) = 1/2$ is equivalent to showing:

$$\int_0^\infty F(\varepsilon_2+\xi)F(-\varepsilon_2+\xi)dF(\varepsilon_2) = \int\int_{\infty}^{\varepsilon_2+\xi} F(\varepsilon_1 - 2\varepsilon_2)dF(\varepsilon_1)dF(\varepsilon_2) \tag{3}$$

To demonstrate that, let $r(\varepsilon_2) \equiv \int_{-\infty}^{\varepsilon_2+\xi} F(\varepsilon_1 - 2\varepsilon_2)dF(\varepsilon_1)$. Integrating by parts we obtain:

$$r(-\varepsilon_2) = F(\varepsilon_2+\xi)F(-\varepsilon_2+\xi) - r(\varepsilon_2)$$

Now, with the help of a change of variables, the right-hand side of (3) can be written as:

$$\begin{aligned}\int_{-\infty}^{0} r(\varepsilon_2)dF(\varepsilon_2) + \int_{0}^{\infty} r(\varepsilon_2)dF(\varepsilon_2) &= \int_{0}^{\infty} r(-\tilde{\varepsilon}_2)dF(\tilde{\varepsilon}_2) + \int_{0}^{\infty} r(\varepsilon_2)dF(\varepsilon_2) \\ &= \int_{0}^{\infty} [r(-\varepsilon_2) + r(\varepsilon_2)]dF(\varepsilon_2) = \int_{0}^{\infty} F(\varepsilon_2 + \xi)F(-\varepsilon_2 + \xi)dF(\varepsilon_2)\end{aligned}$$

which concludes the proof of (3).

(ii) It is enough to note that if T is linear $P(Y_k - 2Y_j + Y_i < 0|\Xi) = P(\varepsilon_{kj} < \varepsilon_{ji}|\Xi)$, which implies $\tilde{\theta} = 0$.

(iii) This part follows from an argument mirroring the one in (i). □

Remark 1. *If we defined our test statistic using two independent pairs of observations, namely:*

$$\tilde{S}_n = \frac{\sqrt{n}}{n(n-1)(n-2)(n-3)} \sum_{i \neq j \neq k \neq l} \mathbb{1}\{Y_i < Y_j < Y_k < Y_m\} sgn(Y_m - Y_k - Y_j + Y_i) K_h \left((X_{mk} - X_{kj})'\hat{\beta}\right)$$

the requirement for symmetry of the distribution of ε could potentially be dropped. This would however come at the increased computational cost as we have a 4-th order U statistic now. Also as we require four observations on Y with approximately equally spaced index values $X'\beta$ instead of triples in our original statistic, the test based on $\tilde{S}_n$ is likely to have lower finite sample power than our baseline test.

In order to obtain critical value for our tests we will assume that the model under the null hypothesis is linear. This is the worst-case H_0 for testing concavity/convexity as any small local deviation from linearity violates the hypothesis. This can also be seen from the proof of Proposition 1. In order to derive the asymptotic distribution of our statistic we make the following assumptions.

Assumption 1. *(a) The kernel $K(\cdot)$ is a bounded, nonnegative, symmetric, twice continuously differentiable function with support on $[-1, 1]$ and uniformly bounded derivatives satisfying:*

(i) $\int K(s)ds = 1$

(ii) $\int s^2 K(s)ds < \infty$

(b) $h \to 0$ *and* $nh^4/\log^4 n \to \infty$ *as* $n \to \infty$

(c) Conditional on the remaining regressors, the distribution of the first element of X is absolutely continuous with respect to the Lebesgue measure, with bounded and twice continuously differentiable density and uniformly bounded second derivatives. The matrix $E[X_i X_i']$ has bounded elements.

(d) The density of ε is bounded and twice continuously differentiable, the derivatives are uniformly bounded.

(e) The estimator of β satisfies: $\hat{\beta} - \beta = \frac{1}{n}\sum_{i=1}^{n} \Omega_i + o_p(n^{-1/2})$, with $E[\Omega_i] = 0$ and bounded $E[\Omega_i \Omega_i']$.

The bandwidth rate condition in Assumption 1(b) is rather weak and allows standard "rule-of-thumb" bandwidth choice $h \sim n^{-1/5}$. Assumption 1(e) is satisfied by a wide range of estimators including the OLS estimator in the linear model, Han's MRC or Ichimura's semiparametric least squares (see e.g. appendix in Szydłowski (2020) for discussion).

Let ξ_{ji} denote the index $X_{ji}'\beta$ and $f_{\xi|\xi}$ denote the density of ξ_{kj} given ξ_{ji}. Define:

$$\begin{aligned}
H(Y_i, X_i, \xi_{ji}, \xi_{kj}) &= E[\mathbb{1}\{Y_i < Y_j < Y_k\} sgn(Y_k - 2Y_j + Y_i)|Y_i, X_i, \xi_{ji}, \xi_{kj}] \\
&\quad + E[\mathbb{1}\{Y_j < Y_i < Y_k\} sgn(Y_k - 2Y_i + Y_j)|Y_i, X_i, \xi_{ji}, \xi_{kj}] \\
&\quad + E[\mathbb{1}\{Y_j < Y_k < Y_i\} sgn(Y_i - 2Y_k + Y_j)|Y_i, X_i, \xi_{ji}, \xi_{kj}] \\
\delta(Y_i, X_i, \xi_{ji}, \xi_{kj}) &= H(Y_i, X_i, \xi_{ji}, \xi_{kj}) f_{\xi|\xi}(\xi_{kj}|\xi_{ji} = (X_j - X_i)'\beta) \\
G(\xi_{ji}, \xi_{kj}) &= E[\mathbb{1}\{Y_i < Y_j < Y_k\} sgn(Y_k - 2Y_j + Y_i)(X_{kj} - X_{ji})'|\xi_{ji}, \xi_{kj}] \\
\mu(\xi_{ji}, \xi_{kj}) &= G(\xi_{ji}, \xi_{kj}) f_{\xi|\xi}(\xi_{kj}|\xi_{ji})
\end{aligned}$$

and let μ_1 denote the derivative of μ with respect to the first argument.

Theorem 1. *If Assumption 1 holds, we have:*

$$\sqrt{n}(U_n - \theta) \to^d N(0, E[\psi_i^2])$$

where $\psi_i = E[\delta(Y_i, X_i, \xi_{ji}, \xi_{kj})|Y_i, X_i] - E[\mu_1(\xi_{ji}, \xi_{ji})]\Omega_i$.

As linearity is the boundary case for testing H_0: $T(\cdot)$ is concave, or H_0: $T(\cdot)$ is convex, we will reject concavity if $S_n < c_\alpha$ and reject convexity if $S_n > c_{1-\alpha}$, where c_α denotes an α quantile from the normal asymptotic distribution. As typical with U-statistics, estimating the variance of the asymptotic distribution of S_n is difficult. On the one hand, the plug-in estimator will involve estimating derivatives of conditional moments and distributions, which requires delicate choices of bandwidths and, essentially, calculation of higher order U-statistics. On the other hand, using the approach in Ghosal et al. (2000) will lead to a 5-th order U-statistic, which would be difficult to calculate with sample sizes typically encountered in applications.[1] Thus, in Section 4 we resort to bootstrap for calculating the critical value as bootstrapping involves only repeated calculation of a 3-rd order U-statistic, U_n, which we find computationally easier than the aforementioned methods.

The global test introduced in this section only has power against global deviations from concavity/linearity/convexity and does not have power if the function is both convex and concave on different parts of the domain. Thus, it can be used as a first check – for example, if the test rejects concavity the researcher concludes that the function cannot be globally concave. Failure to reject would, then, mean that one has to consider our local test described in the next section, in order to verify if indeed the function is globally concave.

3.2 Local test

The main idea of the local test is to consider only triples of the kind portrayed in Figure 2 local to a point y in the domain of the transformation function. In other words, we will check if the transformation function is concave/linear/convex locally around y. Local concavity may be of interest by itself or, to verify if the treatment effect is concave globally, we will take the minimum of the local statistics at different points y to run the test. We will focus on the latter as testing local concavity at y can be preformed in a very similar fashion to the global test described in the previous section.

[1] Proceeding as there would involve calculating:

$$\begin{aligned}\hat{\sigma}^2 = \frac{1}{n(n-1)(n-2)(n-3)(n-4)} \sum_{i\neq j\neq k\neq l\neq m} & \left(\mathbb{1}\{Y_i < Y_j < Y_k\} sgn(Y_k - 2Y_j + Y_i) K_h\left(X'_{kj}\hat{\beta} - X'_{ji}\hat{\beta}\right) + \tilde{\mu}_1 \Omega_i\right) \\ & \times \left(\mathbb{1}\{Y_i < Y_l < Y_m\} sgn(Y_m - 2Y_l + Y_i) K_h\left(X'_{ml}\hat{\beta} - X'_{li}\hat{\beta}\right) + \tilde{\mu}_1 \Omega_i\right) \\ & + \text{symmetric terms,}\end{aligned}$$

where $\tilde{\mu}_1$ is an estimator of $-3E[\mu_1(\xi_{ji}, \xi_{ji})]$ (which itself is a 3-rd order U statistic).

Formally, define:

$$U_n(y) = \frac{1}{n(n-1)(n-2)} \sum_{i \neq j \neq k} \mathbb{1}\{Y_i < Y_j < Y_k\} sgn(Y_k - 2Y_j + Y_i) K_h(Y_i - y) K_h(Y_j - y) \\ \times K_h(Y_k - y) K_h\left((X_{kj} - X_{ji})'\hat{\beta}\right)$$

where in practice the bandwidth used for $(Y_i - y)$ can be different than for $(X_{kj} - X_{ji})'\hat{\beta}$ but, in order to simplify exposition and mathematical arguments, we assume that both bandwidths are of the same order and denote both by h. Now our local test statistic for testing concavity is defined as:

$$S_n^{conc} = \inf_{y \in \mathcal{Y}} \sqrt{n} U_n(y)$$

and the statistic for testing convexity is defined with sup replacing inf above. Linearity can be tested with either of these statistics. For the rest of the article we concentrate on S_n^{conc} as results for testing convexity and linearity follow by very similar arguments.

Intuitively, low values of S_n^{conc} show that there is a large deviation from concavity at some point y and, hence, the treatment effects are not accelerating on the whole domain of the outcome.[2] Therefore, the null hypothesis of concavity would be rejected if $S_n^{conc} < c_\alpha^{conc}$ where c_α^{conc} is an appropriate quantile from the asymptotic approximation to the distribution of our statistic.

In order to obtain c_α^{conc} one could proceed as in Ghosal et al. (2000): approximate the standardised U-statistic process $\sqrt{n} U_n(y)/\sigma_n(y)$, where $\sigma_n(y)$ is the estimator of the asymptotic variance, by a Gaussian process and then apply the extreme value theory in order to derive the distribution of the infimum. However, this approach would involve estimating $\sigma_n(y)$, which as discussed above is problematic in our setup. Instead we propose to use bootstrap to approximate c_α^{conc}.

Define:

$$\begin{aligned}
\phi_i^I(y) =& E[\mathbb{1}\{Y_i < Y_j < Y_k\} sgn(Y_k - 2Y_j + Y_i) K_h(Y_i - y) K_h(Y_j - y) K_h(Y_k - y) K_h\left((X_{kj} - X_{ji})'\beta\right) | Y_i, X_i] \\
&+ E[\mathbb{1}\{Y_j < Y_i < Y_k\} sgn(Y_k - 2Y_i + Y_j) K_h(Y_i - y) K_h(Y_j - y) K_h(Y_k - y) K_h\left((X_{ki} - X_{ij})'\beta\right) | Y_i, X_i] \\
&+ E[\mathbb{1}\{Y_j < Y_k < Y_i\} sgn(Y_i - 2Y_k + Y_j) K_h(Y_i - y) K_h(Y_j - y) K_h(Y_k - y) K_h\left((X_{ik} - X_{kj})'\beta\right) | Y_i, X_i] \\
\phi_i^{II}(y) =& E[\mathbb{1}\{Y_i < Y_j < Y_k\} sgn(Y_k - 2Y_j + Y_i) K_h(Y_i - y) K_h(Y_j - y) K_h(Y_k - y) K_h'\left((X_{kj} - X_{ji})'\beta\right) (X_{kj} - X_{ji})'] \Omega_i
\end{aligned}$$

[2] Note that by the formulas in Example 1 concavity of T is equivalent to convexity of the outcome in the treatment.

where K_h' denotes the derivative of K_h. The following asymptotic approximation will be useful in justifying our bootstrap procedure.

Theorem 2. *If Assumption 1 holds, then:*

$$\sup_y \left| U_n(y) - \frac{1}{n} \sum_{i=1}^n \phi_i(y) \right| = o_p(n^{-1/2})$$

where $\phi_i(y) = \phi_i^I(y) + \phi_i^{II}(y)$.

4 Bootstrap critical values

Our symmetric wild bootstrap procedure for obtaining the critical value for the global or local test is as follows:

1. Estimate β (e.g. by OLS) and calculate the residuals $\hat{\varepsilon}_i = Y_i - X_i'\hat{\beta}$.

2. Draw a random sample $\{v_i : i = 1, \ldots, n\}$ from a two point distribution on {-1,1} with $P(v_i = -1) = P(v_i = 1) = 1/2$, define $\varepsilon_i^* = v_i \hat{\varepsilon}_i$ and generate Y_i^* by:

$$Y_i^* = X_i'\hat{\beta} + \varepsilon_i^*$$

(alternatively, resample also X_i^* independently and generate $Y_i^* = X_i^{*\prime}\beta + \varepsilon_i^*$).

3. Estimate β (e.g. by OLS) using the bootstrap sample. Let the resulting estimate be denoted by β^*.

4. Calculate the statistic S_n or S_n^{conc} on the bootstrap sample using β^* instead of $\hat{\beta}$. Denote the resulting bootstrap statistics by S_n^* and $S_n^{conc,*}$.

5. Obtain the empirical distribution of S_n^* and $S_n^{conc,*}$ by repeating steps 1-4 many times. Calculate the α quantiles of the empirical distribution of S_n^* and $S_n^{conc,*}$ and denote them by c_α^* and $c_\alpha^{conc,*}$, respectively.

Note that step two imposes the null hypothesis of linearity on the bootstrap sample so it "re-centers" the bootstrap statistic on the linear case for which $\theta = 0$. Furthermore, sampling from a symmetric two point distribution preserves symmetry of the error distribution in the bootstrap sample, mirroring the data generating process.

Theorem 3. *If Assumption 1 holds and, additionally, each component of X_i and Ω_i has a finite fourth moment, then:*

$$\lim_{n\to\infty} P\left(\sqrt{n}(U_n - \theta) \leq c^*_\alpha\right) = \alpha.$$

Moreover, if $T(\cdot)$ is linear, we have:

$$\lim_{n\to\infty} P(S_n^{conc} \leq c_\alpha^{conc,*}) = \alpha$$

(and equivalent result holds for a test of linearity or convexity).

The proof of the second result uses Theorem 2.3 in Chernozhukov et al. (2016), which requires the additional finite fourth moment assumption. If β is estimated by OLS, a sufficient condition for a finite fourth moment of Ω_i is that ε_i has a finite fourth moment. Theorem 3 implies, for example, that we can reject global concavity of T, i.e. increasing treatment effect, when $S_n^{conc} < c_\alpha^{conc,*}$, and reject global linearity when $|S_n^{conc}| > c_{1-\alpha/2}^{conc,*}$.

Finally, let us note that the computation of the local statistic involves evaluating a third order U-statistic at different points y and, hence, takes significantly longer to compute than the global statistic. In practice, we recommend running the global test first and then proceed with the local test if the null hypothesis cannot be rejected in the first step.

5 Monte Carlo simulations

The data is generated from the following four models:

$$Y = X + \varepsilon \tag{D0}$$

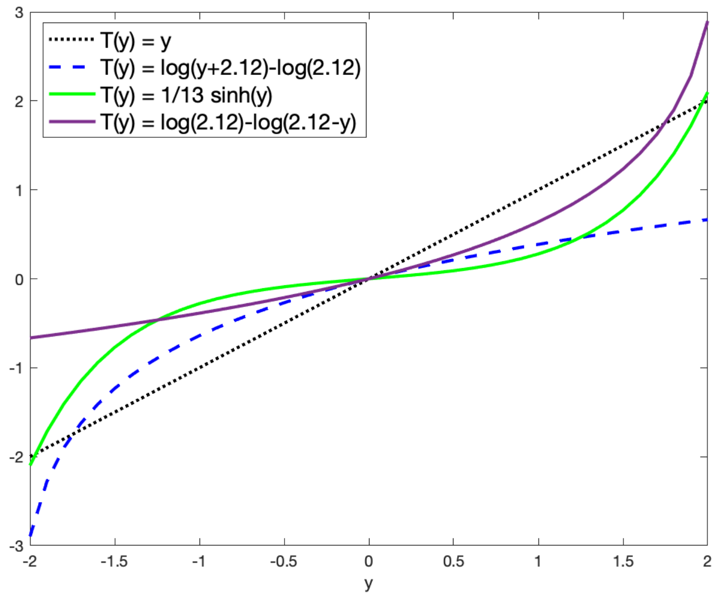
$$\log(Y + 2.12) - \log(2.12) = X + \varepsilon \tag{D1}$$

$$\frac{1}{13}\sinh(2Y) = X + \varepsilon \tag{D2}$$

$$\log(2.12) - \log(2.12 - Y) = X + \varepsilon \tag{D3}$$

where we draw X and ε from the standard normal distribution. Note that D0 is the worst-case model in the null hypothesis, D1 imposes concavity of the transformation, D3 convexity and D2 is neither concave or convex.

Figure 2: Monte Carlo designs



We run 1000 Monte Carlo replications. We use Gaussian kernel functions and rule-of-thumb bandwidths for both $(X_{kj} - X_{ji})'\hat{\beta}$ and Y_i, namely $h = 1.06\hat{\sigma}n^{-1/5}$ where $\hat{\sigma}$ is a sample standard deviation of $(X_{kj} - X_{ji})'\hat{\beta}$ or Y_i. In order to calculate the local statistic S_n^{conc} we use a grid of values for y: -2:0.5:2, and take a minimum over the grid. The number of bootstrap replications used to calculate the critical value is 500 and we consider three sample sizes: $n = 100, 250$ and 500.

Table 1: Test of concavity, rejection probabilities, 5% level

		Global test			Local test		
		$n = 100$	$n = 250$	$n = 500$	$n = 100$	$n = 250$	$n = 500$
H0 true	D0	0.072	0.057	0.050	0.055	0.042	0.039
H0 true	D1	0.000	0.000	0.000	0.000	0.000	0.000
H0 false	D2	0.093	0.090	0.093	0.499	0.854	0.953
H0 false	D3	1.000	1.000	1.000	1.000	1.000	1.000

Note: 1000 Monte Carlo simulations, 500 bootstrap replications.

Table 1 contains the results of the Monte Carlo simulations. We concentrate on testing concavity, as the results for linearity and convexity tests, are very similar. The rejection probabilities in the linear case (D0) are close to the nominal level for both the global and local test and both test have perfect power against a globally convex alternative in D3.

As predicted, the global test has low power against D2, for which the transformation function is concave on half of the domain and convex on the other half. In this case deviations from concavity, as measured by our global statistic, cancel with positive values of the statistic obtained for the part of the domain where the function is concave, resulting in the global statistic taking values close to zero, just as for the linear case.

For design D2, our local test significantly improves over the global test with almost perfect detection of non-concavity for $n = 500$. This is in line with the intuition that the local test will concentrate on the region of largest violation of concavity instead of averaging over the measures of concavity for different regions. Overall, Table 1 shows very good performance of both of our procedures.

6 Application: Curvature of loan demand

We use data from Karlan & Zinman (2008), who ran randomised trials with a for-profit consumer lender in South Africa targeting high-risk consumer loan market. The lender randomised individual interest rate direct mail offers to over 50,000 former clients, conditional on the client's risk category. Karlan & Zinman (2008) found that the loan demand curves are downward sloping. We will investigate if the demand curves are convex or, in other words, if increases in interest rate have a stronger negative effect on loan demand the lower the rate.

The dependent variable is the amount borrowed (in rands) at the offered interest rate.[3] The experiment was ran in three mailer waves over four months, thus as in Karlan & Zinman (2008) we include the risk category and wave dummies as controls in X. The data contains 2325 observations. As a first step, we estimated the transformation function in our model using the estimator in Chen (2002) and plotted it in Figure 3 together with a smoothed version.[4] The figure suggests that the transformation is not far from being concave on most of its domain, maybe besides the values of loan size below 1000 rands, which implies convex demand curve by formulas in Example 1.

[3] We abstract from selectivity issues here. See Karlan & Zinman (2008) for discussion.

[4] This is for illustrative purposes only. Note that the estimator in Chen (2002) does not impose monotonicity, so a full exercise of estimating T should include additional step of monotonising the estimate. We want to stress that our testing procedure does not rely on any estimator of T.

Figure 3: Estimated transformation

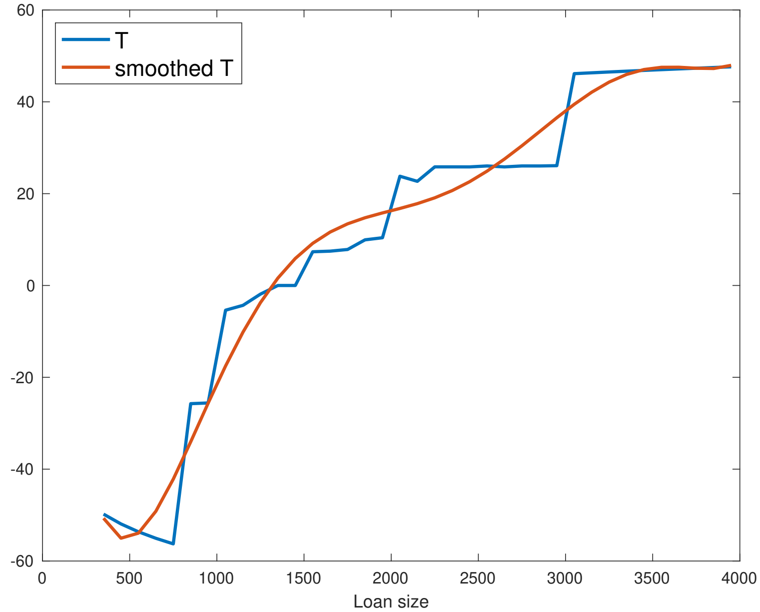


Table 2: Testing curvature of loan demand, 5% level

H_0	Global test		Local test		Local test, $loan \geq 1000$	
	Statistic	Reject H_0?	Statistic	Reject H_0?	Statistic	Reject H_0?
convexity	14.11	No	-4.42	Yes	-2.64	No
linearity	14.11	Yes				
concavity	14.11	Yes				

Data source: Karlan & Zinman (2008), replication files

In order to formally test our conjectures, we first apply our global test to verify if the demand function is concave, linear or convex. As Table 2 shows, global test rejects linearity and concavity of demand. Thus, in the second step we apply the local test to the null hypothesis of convexity and, in fact, reject the null hypothesis, concluding that the loan demand is not globally convex in the interest rate. To shed more light on where the non-convexity may come from, we re-ran the test on the sample excluding loans lower than 1000 rands and found that convexity is not rejected on this sub-sample. Therefore, overall we confirm that loan demand is mostly convex in interest rate, besides very small loans sector of the market.

7 Conclusion

Our application demonstrates usefulness of our testing procedures in recovering monotonicity of treatment effects. Particular appeal of the procedures described in this article comes from the fact that they avoid estimation of the transformation function and only require OLS estimation of the vector of coefficients β. Thus, they are easy to implement. Additionally, computation of the third order U-statistic involved in our tests can be done efficiently by sorting the data by Y first – this reduces computational complexity to $O(nlog(n) + n(n-2)/2)$ from $O(n^3)$ for a straightforward triple loop through the observations.

Appendix

A Proofs

Let $\mathcal{G} = \{g_y(w_1, w_2, \ldots, w_m) : y \in \mathcal{Y} \subset \mathbb{R}\}$ be a family of symmetric, real-valued functions defined on $\mathcal{W}^m$. We will use the operator notation common in the U-statistics literature. For example, for the case of $m = 2$ we will have $P^0 h = h$, $P^2 h = \int\int g(w_1, w_2) dP(w_1) dP(w_2)$, $P_n g(w_1) = 1/n \sum_{i=1}^n g(w_1, W_i)$ and $P_n^* g(w_1) = 1/n \sum_{i=1}^n g(w_1, W_i^*)$ etc. We say that a symmetric function g is P-canonical if $Pg(w_1, \ldots, w_{m-1}, \cdot) = 0$ for almost all $w_1, \ldots, w_{m-1}$.

Let $\|\cdot\|_{\mathcal{G}} \equiv \sup_{g \in \mathcal{G}} \|\cdot\|$ where $\|\cdot\|$ is the Euclidean norm and $\|\cdot\|_\infty, \|\cdot\|_{P,q}$ denote the sup and the $L_q(P)$ norm, respectively. Define a U-process:

$$U_n^{(m)} g = \frac{(n-m)!}{n!} \sum_{i_1, i_2, \ldots, i_m \text{ distinct}} g(W_{i_1}, W_{i_2}, \ldots, W_{i_m})$$

and denote the same process evaluated on a bootstrap sample as $U_n^{*(m)} g$. Furthermore, define

$$\pi_{k,m}^P g(w_1, \ldots, w_k) = (\delta_{w_1} - P) \ldots (\delta_{w_k} - P) P^{m-k} g$$

where $\delta_{w_1} g = g(w_1, \cdot)$.

We will frequently use the following stochastic order arithmetic, for a sequence a_n:

$$o_p^*(a_n) + o_p(a_n) = o_p(a_n), \quad O_p^*(a_n) + O_p(a_n) = O_p(a_n)$$

which follows from the Law of Iterated Expectations. See Szydłowski (2020) for more discussion. We will also write $\lesssim$ for inequality up to a multiplicative constant where the constant does not depend on the sample size n or the sample data (but may depend on m and characteristics of $\mathcal{G}$).

A.1 Useful lemmas

Lemma 1. *Let $\mathcal{G}$ be a class of P-canonical, Euclidean functions with envelope G satisfying $P^m G^2 < \infty$. Then:*

$$P\|U_n^{*(m)}g\|_{\mathcal{G}} \lesssim n^{-m/2}\sqrt{(P^m G^2)^\alpha}$$

where $0 < \alpha < 1$ depends only on m and characteristics of $\mathcal{G}$.

Proof. It follows from the proof of Lemma 2(c) in Szydłowski (2020) that:

$$P\|U_n^{*(m)}g\|_{\mathcal{G}} \lesssim P\|U_n^{*(m)}\tilde{g}\|_{\tilde{\mathcal{G}}}$$

where $\tilde{g}((N_{i_1}^{(1)}, W_{i_1}), \ldots, (N_{i_m}^{(m)}, W_{i_m})) = N_{i_1}^{(1)} \ldots N_{i_m}^{(m)} g(W_{i_1}, \ldots, W_{i_m})$ which is contained in the class $\tilde{\mathcal{G}}$, and $\{N_i^{(k)} : i = 1, \ldots, n\}_{k=1}^m$ denote independent copies of a sequence of independent Poisson random variables with parameter $1/2$. Note that the class of functions $\tilde{\mathcal{G}}$ inherits its properties from $\mathcal{G}$, in particular it is a Euclidean class with envelope $\tilde{G}$ satisfying $P^m \tilde{G}^2 < \infty$.

Now the Main Corollary and argument in the proof of Corollary 4 in Sherman (1994) together with Cauchy-Schwartz inequality give:

$$P\|U_n^{*(m)}\tilde{g}\|_{\tilde{\mathcal{G}}} \lesssim n^{-m/2}\sqrt{(P^m \tilde{G}^2)^\alpha} \lesssim n^{-m/2}\sqrt{(P^m G^2)^\alpha}$$

which concludes the proof. □

Lemma 2. *Let $\mathcal{G}$ be a Euclidean class of functions. We have:*

$$P\left(\left\|U_n^{(m)}g - \sum_{k=0}^{r-1} U_n^{(k)}\pi_{k,m}^P g\right\|_{\mathcal{G}}\right) \lesssim n^{-r/2}\sqrt{P^m G^2}$$
$$P\left(\left\|U_n^{(m)}g - P^m g\right\|_{\mathcal{G}}\right) \lesssim n^{-1/2}\sqrt{P^m G^2}$$

Proof. From Theorem A.1 in Ghosal et al. (2000) and the discussion that follows we have:

$$P\left(\left\|U_n^{(m)}g - \sum_{k=0}^{r-1} U_n^{(k)}\pi_{k,m}^P g\right\|_{\mathcal{G}}\right) \lesssim n^{-r/2}\sqrt{P^m G^2}\int_0^1 \sup_{Q:Q \text{ discrete}} \log N(\varepsilon\|G\|_{Q,2}, \mathcal{G}, L_2(Q))^{r/2} d\varepsilon$$

where $N(\cdot)$ denotes a covering number as in Definition 2.1.5 in Van der Vaart & Wellner (1996). The integral on the right-hand side is bounded as $\mathcal{G}$ is Euclidean (ibid., Ch.2).

The second part follows from Theorem A.2 in Ghosal et al. (2000) and the same reasoning as above. $\square$

Lemma 3. *Assume that:*

(a) The functional $B : \mathcal{G} \to \mathbb{R}$ satisfies: there exists a countable subset $\bar{\mathcal{G}}$ of $\mathcal{G}$ such that for any $g \in \mathcal{G}$, there exists a sequence $\bar{g}_m \in \bar{\mathcal{G}}$ with $\bar{g}_m \to g$ pointwise and $B(\bar{g}_m) \to B(g)$.

(b) The class of functions $\mathcal{G}$ is Euclidean with a measurable envelope G.

(c) There exist constants $b \geq \sigma > 0$ and $q \in [4, \infty)$ such that $\sup_{g\in\mathcal{G}} P|g|^k \leq \sigma^2 b^{k-2}$ for $k = 2, 3, 4$ and $\|G\|_{P,q} \leq b$.

Let $\mathbb{G}_n = \sqrt{n}P_n$, $\mathbb{G}^*_n = \sqrt{n}P^*_n$ and $\mathbb{G}_P$ denote a centred Gaussian process indexed by $\mathcal{G}$ with covariance function $E[\mathbb{G}_P f \times \mathbb{G}_P g] = Cov(f(W), g(W))$. Let $N_B(\varepsilon)$ denote the bracketing number of the class of functions $\mathcal{BG} = \{B(g) : g \in \mathcal{G}\}$.

Then for every $\gamma \in (0, 1)$ there exists random variables Z and Z^* which follow the same distribution as $\|B(g) + \mathbb{G}_P g\|_{\mathcal{G}}$ (the latter, conditionally on the sample), such that:

$$
\begin{aligned}
|\|B(g) + \mathbb{G}_n g\|_{\mathcal{G}} - Z| &= O_p\left(\frac{bK_n}{\gamma^{1/2}n^{1/2-1/q}} + \frac{(b\sigma^2K_n^2)^{1/3}}{\gamma^{1/3}n^{1/6}} + \frac{(b\sigma K_n^{3/2})^{1/2}}{\gamma^{1/2}n^{1/4}}\right)\\
|\|B(g) + \mathbb{G}^*_n g\|_{\mathcal{G}} - Z^*| &= O_p\left(\frac{bK_n}{\gamma^{1+1/q}n^{1/2-1/q}} + \frac{(b\sigma^2K_n^2)^{1/3}}{\gamma^{1/3}n^{1/6}} + \frac{(b\sigma K_n^{3/2})^{1/2}}{\gamma^{1+1/q}n^{1/4}}\right)
\end{aligned}
$$

where K_n is of order $\log N_B(\varepsilon) + \log n \vee \log(b/\sigma)$ and $K_n^3 \leq n$.

Proof. This result follows directly from Corollary 2.2 in Chernozhukov et al. (2014) and Theorem 2.3 in Chernozhukov et al. (2016). Note that the differences in rates in the sample and bootstrap version are usually only of $\log n$ order so they will not be essential to our rate results. $\square$

Lemma 4. *Let K be a symmetric kernel function supported on $[-1, 1]$. We have:*

$$\int\int\int \mathbb{1}\{s_1 < s_2 < s_3\} sgn(s_3 - 2s_2 + s_1)K(s_1)K(s_2)K(s_3)ds_1ds_2ds_3 = 0$$

Proof. For simplicity we write $\int_{-1}^{1}\int_{-1}^{1}\int_{-1}^{1}$ as $\int_{-1}^{1}$. We have:

$$\begin{aligned}
&\int_{-1}^{1} \mathbb{1}\{s_1 < s_2 < s_3\} sgn(s_3 - 2s_2 + s_1) K(s_1)K(s_2)K(s_3) ds_1 ds_2 ds_3 \\
&\qquad = \int_{-1}^{1} \mathbb{1}\{s_1 < s_2 < s_3\} sgn(s_3 - 2s_2 + s_1) K(-s_1)K(-s_2)K(-s_3) ds_1 ds_2 ds_3 \\
&\qquad = \int_{-1}^{1} \mathbb{1}\{\tilde{s}_1 > \tilde{s}_2 > \tilde{s}_3\} sgn(-\tilde{s}_3 + 2\tilde{s}_2 - \tilde{s}_1) K(\tilde{s}_1)K(\tilde{s}_2)K(\tilde{s}_3) d\tilde{s}_1 d\tilde{s}_2 d\tilde{s}_3 \\
&\qquad = -\int_{-1}^{1} \mathbb{1}\{\tilde{s}_3 < \tilde{s}_2 < \tilde{s}_1\} sgn(\tilde{s}_1 - 2\tilde{s}_2 + \tilde{s}_3) K(\tilde{s}_1)K(\tilde{s}_2)K(\tilde{s}_3) d\tilde{s}_1 d\tilde{s}_2 d\tilde{s}_3
\end{aligned}$$

where the first equality follows from symmetry of K, second from change of variables and the third from $sgn(-x) = -sgn(x)$. Finally note that the last integral is equal to the initial one after renaming the variables, so the result follows. $\square$

Lemma 5. *Let K satisfy Assumption 1(a). The classes of functions:*

$$\begin{aligned}
\mathcal{F}_1 &= \left\{ \mathbb{1}\{y_1 < y_2 < y_3\} sgn(y_3 - 2y_2 + y_1) K\left(\frac{y_1 - y}{h}\right) K\left(\frac{y_2 - y}{h}\right) K\left(\frac{y_3 - y}{h}\right) K\left(\frac{(x_{32} - x_{21})'\beta}{h}\right) : y \in \mathcal{Y} \right\} \\
\mathcal{F}_2 &= \Bigg\{ \mathbb{1}\{y_1 < y_2 < y_3\} sgn(y_3 - 2y_2 + y_1) K\left(\frac{y_1 - y}{h}\right) K\left(\frac{y_2 - y}{h}\right) K\left(\frac{y_3 - y}{h}\right) K'\left(\frac{(x_{32} - x_{21})'\beta}{h}\right) \\
&\qquad \times (x_{32} - x_{21})' : y \in \mathcal{Y} \Bigg\} \\
\mathcal{F}_3 &= \Bigg\{ \mathbb{1}\{y_1 < y_2 < y_3\} sgn(y_3 - 2y_2 + y_1) K\left(\frac{y_1 - y}{h}\right) K\left(\frac{y_2 - y}{h}\right) K\left(\frac{y_3 - y}{h}\right) K''\left(\frac{(x_{32} - x_{21})'\beta}{h}\right) \\
&\qquad \times (x_{32} - x_{21})(x_{32} - x_{21})' : y \in \mathcal{Y} \Bigg\}
\end{aligned}$$

are Euclidean with envelopes $F_1 = \|K\|_\infty^4 \mathbb{1}\{|y_1 - y_2| < 2h\}\mathbb{1}\{|y_1 - y_3| < 2h\}\mathbb{1}\{|y_2 - y_3| < 2h\}\mathbb{1}\{|(x_{32} - x_{21})'\beta| < h\}$, $F_2 = F_1\|K'\|_\infty/\|K\|_\infty(x_{32} - x_{21})'$ and $F_3 = F_1\|K''\|_\infty/\|K\|_\infty(x_{32} - x_{21})(x_{32} - x_{21})'$, respectively.

Proof. Note that $\mathcal{F}_1$, $\mathcal{F}_2$ and $\mathcal{F}_3$ are products of single functions and classes of functions of the form $K\left(\frac{\cdot - y}{h}\right)$, $K'\left(\frac{\cdot - y}{h}\right)$ or $K''\left(\frac{\cdot - y}{h}\right)$, which are Euclidean by Example 2.10 in Pakes & Pollard (1989). As the product of Euclidean classes is Euclidean (see Lemma 2.14 (ibid.)), the result follows. The form of the envelope follows from the fact that K is supported on $[-1, 1]$. $\square$

B Proof of Theorem 1

To simplify notation set $\theta = 0$. The argument for $\theta \neq 0$ follows *verbatim* by recentering the test statistic appropriately. Write U_n as:

$$\begin{aligned}
U_n = & \frac{1}{n(n-1)(n-2)} \Big\{ \sum_{i \neq j \neq k} \mathbb{1}\{Y_i < Y_j < Y_k\} sgn(Y_k - 2Y_j + Y_i) K_h \left((X_{kj} - X_{ji})'\beta\right) && \text{(I)} \\
& + \sum_{i \neq j \neq k} \mathbb{1}\{Y_i < Y_j < Y_k\} sgn(Y_k - 2Y_j + Y_i) K_h' \left((X_{kj} - X_{ji})'\beta\right) (X_{kj} - X_{ji})'(\hat{\beta} - \beta) && \text{(II)} \\
& + \sum_{i \neq j \neq k} \mathbb{1}\{Y_i < Y_j < Y_k\} sgn(Y_k - 2Y_j + Y_i) K_h'' \left((X_{kj} - X_{ji})'\tilde{\beta}\right) (\hat{\beta} - \beta)'(X_{kj} - X_{ji})(X_{kj} - X_{ji})'(\hat{\beta} - \beta) \Big\} && \text{(III)}
\end{aligned}$$

where $\tilde{\beta}$ is between β and $\hat{\beta}$.

Note that by Cauchy-Schwartz inequality the third term in this decomposition can be bounded as:

$$|(III)| \leq h^{-3} \|K''\|_\infty \|\hat{\beta} - \beta\|^2 \left\| \frac{1}{n(n-1)(n-2)} \sum_{i \neq j \neq k} \mathbb{1}\{Y_i < Y_j < Y_k\} sgn(Y_k - 2Y_j + Y_i)(X_{kj} - X_{ji})(X_{kj} - X_{ji})' \right\|$$

Now $\|\hat{\beta} - \beta\|^2 = O_p(n^{-1})$ by Assumption 1(e) and the U-statistic on the right-hand side is $O_p(n^{-1/2})$ by Theorem 5.5.1A in Serfling (1980), which gives:

$$(III) = O_p(n^{-3/2} h^{-3}) = o_p(n^{-1/2})$$

under our bandwidth conditions.

Further, using Assumption 1(e) I can rewrite (II) as:

$$\begin{aligned}
\frac{1}{n(n-1)(n-2)} & \sum_{i \neq j \neq k} \mathbb{1}\{Y_i < Y_j < Y_k\} sgn(Y_k - 2Y_j + Y_i) K_h' \left((X_{kj} - X_{ji})'\beta\right) (X_{kj} - X_{ji})' P_n \Omega = \\
& = \frac{1}{n(n-1)(n-2)(n-3)} \sum_{i \neq j \neq k \neq l} \mathbb{1}\{Y_i < Y_j < Y_k\} sgn(Y_k - 2Y_j + Y_i) K_h' \left((X_{kj} - X_{ji})'\beta\right) \\
& \quad \times (X_{kj} - X_{ji})' \Omega_l + o_p(n^{-1/2})
\end{aligned}$$

where the equality follows from Lemma 5.7.3 in Serfling (1980).

It remains to apply Lemma 2 to (I) and (II) with $r = 2$ and $\mathcal{G}$ containing a single function. In order to do that, let g^I and g^{II} be symmetrised kernels of the U-statistics in (I) and (II). Note that $(I) = U_n^3 g^I$ and $(II) = U_n^3 g^{II}$ and the envelopes satisfy:

$$E(G^I)^2 \lesssim h^{-2} E[\mathbb{1}\{|y_1 - y_2| < 2h\}\mathbb{1}\{|y_1 - y_3| < 2h\}\mathbb{1}\{|y_2 - y_3| < 2h\}\mathbb{1}\{|(x_{32} - x_{21})'\beta| < h\}] = O(h)$$

$$E(G^{II,'}G^{II}) \lesssim h^{-4} E[\mathbb{1}\{|y_1 - y_2| < 2h\}\mathbb{1}\{|y_1 - y_3| < 2h\}\mathbb{1}\{|y_2 - y_3| < 2h\}\mathbb{1}\{|(x_{32} - x_{21})'\beta| < h\} \times (x_{32} - x_{21})'(x_{32} - x_{21})] = O(h^{-1})$$

which, using Lemma 2 gives:

$$P\left(\left\|U_n^{(m)}(g^I + g^{II}) - P_n \pi_{1,3}^P (g^I + g^{II})\right\|\right) = O(n^{-1}h^{-1/2}) = o_p(n^{-1/2})$$

under our bandwidth conditions in Assumption 1(b).

Now by direct calculation, using change of variables, Taylor expansion and $\int K'(u)du = 0, \int K'(u)udu = -1$ we get $\pi_{1,3}^P(g^I + g^{II}) = E[\delta(Y_i, X_i, \xi_{ji}, \xi_{kj})|Y_i, X_i] - E[\mu_1(\xi_{ji}, \xi_{ji})]\Omega_i$ and the result follows from the CLT.

C Proof of Theorem 2

Let $K_h(\mathbf{Y} - y) \equiv K_h(Y_i - y)K_h(Y_j - y)K_h(Y_k - y)$. Write $U_n(y)$ as:

$$\frac{1}{n(n-1)(n-2)}\Bigg\{ \sum_{i \neq j \neq k} \mathbb{1}\{Y_i < Y_j < Y_k\} sgn(Y_k - 2Y_j + Y_i) K_h(\mathbf{Y} - y) K_h\left((X_{kj} - X_{ji})'\beta\right) \quad \text{(I)}$$

$$+ \sum_{i \neq j \neq k} \mathbb{1}\{Y_i < Y_j < Y_k\} sgn(Y_k - 2Y_j + Y_i) K_h(\mathbf{Y} - y) K_h'\left((X_{kj} - X_{ji})'\beta\right)(X_{kj} - X_{ji})'(\hat\beta - \beta) \quad \text{(II)}$$

$$+ (\hat\beta - \beta)' \sum_{i \neq j \neq k} \mathbb{1}\{Y_i < Y_j < Y_k\} sgn(Y_k - 2Y_j + Y_i) K_h(\mathbf{Y} - y) K_h''\left((X_{kj} - X_{ji})'\tilde\beta\right)(X_{kj} - X_{ji})(X_{kj} - X_{ji})'(\hat\beta - \beta)\Bigg\} \quad \text{(III)} \qquad (4)$$

where $\tilde\beta$ is between β and $\hat\beta$.

First, we will show that $(III) = o_p(n^{-1/2})$ using the second result in Lemma 2. By standard calculation and Lemma 4:

$$E[\mathbb{1}\{Y_i < Y_j < Y_k\} sgn(Y_k - 2Y_j + Y_i) K_h(\mathbf{Y} - y) K_h''\left((X_{kj} - X_{ji})'\tilde\beta\right)(X_{kj} - X_{ji})(X_{kj} - X_{ji})'] =$$

$$= E\Bigg[\int \mathbb{1}\{s_1 < s_2 < s_3\} sgn(s_3 - 2s_2 + s_1) K(s_1)K(s_2)K(s_3) ds_1 ds_2 ds_3 f(y - X_i'\beta) f(y - X_j'\beta) f(y - X_k'\beta) \times K_h''\left((X_{kj} - X_{ji})'\tilde\beta\right)(X_{kj} - X_{ji})(X_{kj} - X_{ji})'\Bigg] + o(h) = o(h)$$

Let $g_y^{(III)}$ denote the symmetrized version of the function under the expectation above. Note that Lemma 5 applies to the class of functions $\{h^5 g_y^{(III)} : y \in \mathcal{Y}\}$, thus this class is Euclidean. Further note that expectation of the square of the envelope of $g_y^{(III)}$ is of order $O(h^{-7})$ by similar calculation as above. Now Lemma 2 applied to $U_n^{(3)} g_y^{(III)}$ gives $U_n^{(3)} g_y^{(III)} = O_p(n^{-1/2}h^{-7/2}) + o_p(h)$. Thus, we can bound (III):

$$(III) \lesssim \|\hat{\beta} - \beta\|(O_p(n^{-1/2}h^{-7/2}) + o_p(h)) = O_p(n^{-3/2}h^{-7/2}) + o_p(n^{-1}h)$$

and this is $o_p(n^{-1/2})$ under our rate conditions.

Further, let $g_y^{(II)}$ denote the symmetrised version of $\mathbb{1}\{Y_i < Y_j < Y_k\} sgn(Y_k - 2Y_j + Y_i) K_h(\mathbf{Y} - y) K_h' \left((X_{kj} - X_{ji})'\beta\right) (X_{kj} - X_{ji})'\Omega_l$. Again, by Lemma 5.7.3 in Serfling (1980) we have that $(II) = U_n^4 g_y^{(II)} + o_p(n^{-1/2})$.

It remains to apply the first part of Lemma 2 to show that:

$$|(I) + (II) - P^3 g_y^I - P_n(\pi_{1,3}^P g_y^I + \pi_{1,4}^P g_y^{II})| \lesssim O_p(n^{-1}h^{-1/2}) = o_p(n^{-1/2})$$

uniformly over y, and note that $P^3 g_y^I + \pi_{1,3}^P g_y^I = \phi_i^I(y)$ and $\pi_{1,4}^P g_y^{II} = \phi_i^{II}(y)$ with ϕ_i^I, ϕ_i^{II} defined in the main text. We also used $P^4 g_y^{II} = 0$.

D Proof of Theorem 3

The main part of the argument shows that a linear representation equivalent to the one in Theorem 2 holds for the bootstrap statistic. This is done by exploiting an 'in probability' Hoeffding decomposition of the bootstrap statistic in terms of smooth functions, following the idea in Subbotin (2007) (see also Szydłowski (2020)). Then, application of Lemma 3 finishes the proof.

Again, to simplify notation set $E[U_n(y)] = 0$. Consider a decomposition of the form (4) but calculated on the bootstrap sample (i.e. replacing β with β^*, (X_i, Y_i) with (X_i^*, Y_i^*)).[5] Let $(I*), (II*)$ and $(III*)$ denote its elements.

First, we will show $(III*) = o_p(n^{-1/2})$. Note that since β^* is an OLS estimator we have

[5] In the proof we assume that X^* is resampled independently of ε^*. The same argument applies if only ε^* is resampled. Just define $P_n^* = P_{Y,n}^* \times P_X$.

$\beta^* - \beta = \hat{\beta} - \beta + \frac{1}{n}\sum_{i=1}^n \Omega_i^* = O_p(n^{-1/2})$. Let $g_y^{(III)}$ be as defined above. Apply the Hoeffding decomposition in terms of population moments to (III):

$$U_n^{*(3)} g_y^{(III)} = P^3 g_y^{(III)} + P_n^* \pi_{1,3}^P g_y^{(III)} + \sum_{k=2}^{3} \binom{3}{2} U_n^{*(k)} \pi_{k,3}^P g_y^{(III)} \tag{5}$$

Note that we have shown that $P^3 g_y^{(III)} = o(h)$ above. Now, noting that we can always make the envelope of $g_y^{(III)}$ greater than 1 by appropriate scaling and, thus, drop α from Lemma 1 for our purpose, applying Lemma 1 with the help of Lemma 5 we obtain:

$$P \sup_y \|U_n^{*(m)} \pi_{2,3}^P g_y^{(III)}\| \lesssim O(n^{-1}h^{-7/2}) = o_p(n^{-1/2}) \tag{6}$$

$$P \sup_y \|U_n^{*(m)} \pi_{3,3}^P g_y^{(III)}\| \lesssim O(n^{-3/2}h^{-7/2}) = o_p(n^{-1/2}) \tag{7}$$

where we used the fact that envelope of $P^k g_y(w_1, w_2, \ldots, w_m)$ is equal to $P^k G$ where G is the envelope of g_y. This gives $|U_n^{*(3)} g_y^{(III)} - P_n^* \pi_{1,3}^P g_y^{(III)}| = o_p(n^{-1/2}) + o_p(h)$. By Lemma 5 and Lemma A.2 in Ghosal et al. (2000) the class of functions $\bar{\mathcal{G}} = \{\pi_{1,3}^P g_y^{(III)} : y \in \mathcal{Y}\}$ is Glivenko-Cantelli, thus $P_n \pi_{1,3}^P g_y^{(III)} = o_p(1)$ uniformly over y. By Theorem 3.5 in Gine & Zinn (1990) this is equivalent to $\sup_y \|(P_n^* - P_n)\pi_{1,3}^P g_y^{(III)}\| = o_p(1)$. Combining this result, (5),(6) and (7) gives a bound on $(III*)$:

$$|(III*)| \lesssim o_p(n^{-1}) + o_p(n^{-1}h) + o_p(n^{-3/2}) = o_p(n^{-1/2}).$$

Consider $(II*)$ now. We have $\beta^* - \beta = \hat{\beta} - \beta + P_n^* \Omega_i = P_n^* \Omega_i + O_p(n^{-1/2})$. Again, applying Lemma 5.7.3 in Serfling (1980) we can write:

$$(II*) = \frac{1}{n(n-1)(n-2)(n-3)} \sum_{i \neq j \neq k \neq l} \mathbb{1}\{Y_i^* < Y_j^* < Y_k^*\} sgn(Y_k^* - 2Y_j^* + Y_i^*) K_h(\mathbf{Y}^* - y)$$
$$\times K_h'\left((X_{kj}^* - X_{ji}^*)'\beta\right)(X_{kj}^* - X_{ji}^*)'\Omega_l^* + o_p(n^{-1/2})$$

The leading term is equal to $U_n^{*(4)} g_y^{(II)}$ and can be decomposed as :

$$U_n^{*(4)} g_y^{(II)} = P^4 g_y^{(II)} + P_n^* \pi_{1,4}^P g_y^{(II)} + \sum_{k=2}^{4} \binom{4}{2} U_n^{*(k)} \pi_{k,4}^P g_y^{(II)}$$

Note that $P^4 g_y^{(II)} = 0$ as $E[\Omega_i] = 0$ by Assumption 1(e). Furthermore, the envelopes of $\pi_{k,4} g_y^{(II)}$, $k = 2, 3, 4$ are given by $P^{4-k} F_2 \Omega_i$ where F_2 is defined in Lemma 5 and, under Assumption 1, satisfy $P^2(P^2 F_2 \Omega_i)^2 = O(h^{-4})$, $P^3(P F_2 \Omega_i)^2 = O(h^{-5})$ and $P^4(F_2 \Omega_i)^2 = O(h^{-7})$. Now applying Lemma 1 we obtain:

$$
\begin{aligned}
P \sup_y |U_n^{*(2)} \pi_{2,4}^P g_y^{(II)}| &= O(n^{-1} h^{-2}) \\
P \sup_y |U_n^{*(3)} \pi_{3,4}^P g_y^{(II)}| &= O(n^{-3/2} h^{-5/2}) \\
P \sup_y |U_n^{*(4)} \pi_{4,4}^P g_y^{(II)}| &= O(n^{-2} h^{-7/2})
\end{aligned}
$$

which implies $|U_n^{*(4)} g_y^{(II)} - P_n^* \pi_{1,4}^P g_y^{(II)}| = o_p(n^{-1/2})$ uniformly over y.

Consider $(I*)$. We can decompose:

$$
(I*) = P^3 g_y^{(I)} + P_n^* \pi_{1,3}^P g_y^{(I)} + \sum_{k=2}^{3} \binom{3}{2} U_n^{*(k)} \pi_{k,3}^P g_y^{(I)}
$$

Using Lemmas 1 and 5 and a similar reasoning as for $(II*)$ one can show that:

$$
P \sup_y |U_n^{*(2)} \pi_{2,3}^P g_y^{(I)}| = O(n^{-1} h^{-3/2}) \quad P \sup_y |U_n^{*(3)} \pi_{3,3}^P g_y^{(I)}| = O(n^{-3/2} h^{-2})
$$

which implies $|U_n^{*(3)} g_y^{(I)} - P^3 g_y^{(I)} - P_n^* \pi_{1,3}^P g_y^{(I)}| = o_p(n^{-1/2})$ uniformly over y.

Recall that $S_n^{conc,*} = \sup_y \sqrt{n}[(I*) + (II*) + (III*)] = sup_y \sqrt{n} P_n^* (P^3 g_y^{(I)} + \pi_{1,3}^P g_y^{(I)} + \pi_{1,4}^P g_y^{(II)}) + o_p(1) = sup_y \{\sqrt{n} P^3 g_y^{(I)} + \sqrt{n} P_n^* (\phi_i^{(Ic)}(y) + \phi_i^{(II)}(y))\} + o_p(1)$, where $\phi_i^{(Ic)}(y) \equiv \phi_i^{(I)}(y) - P^3 g_y^{(I)}$. Denote $B(y) \equiv \sqrt{n} P^3 g_y^{(I)}$ and $\phi_i^c(y) \equiv \phi_i^{(Ic)}(y) + \phi_i^{(II)}(y)$ and observe that by Lemma 5 the class $\{g_y^{(I)} : y \in \mathcal{Y}\}$ is Euclidean, which implies $N_B(\varepsilon) \sim \log n$ where $N_B(\varepsilon)$ is defined in Lemma 3. We will now show that $|S_n^{conc,*} - \sup_y \{B(y) + \mathbb{G}_P \phi_i^c(y)\}| = o_p(1)$ where $\mathbb{G}_P \phi_i^c(y)$ is a centred Gaussian process with covariance function $\rho(y_1, y_2) = Cov(\phi_i^c(y_1), \phi_i^c(y_2))$.

In order to demonstrate that, let us verify the conditions of Lemma 3. Firstly, condition (a) is satisfied, for example, by taking $\bar{\mathcal{G}} = \{P^3 g_y^{(I)} : y \in \mathcal{Y}_r\}$ where $\mathcal{Y}_r$ denotes all rational numbers within the interval $\mathcal{Y}$ and by noting that $P^3 g_y^{(I)}$ is continuous under our assumptions. Condition (b) follows from Lemma 5 and Lemma A.2 in Ghosal et al. (2000). Next, in order to show that condition (c) of Lemma 3 is satisfied, it is enough to verify it for $\tilde{\phi}_i^{(I)}(y) \equiv h \phi_i^{(I)}(y)$ and

$\tilde{\phi}_i^{(II)}(y) \equiv h\phi_i^{(II)}(y)$. Let $G_{h,y}(Y_i, X_i) \equiv E[\mathbb{1}\{Y_i < Y_j < Y_k\} sgn(Y_k - 2Y_j + Y_i) K_h(Y_j - y) K_h(Y_k - y) K_h((X_{kj} - X_{ji})'\beta) | Y_i, X_i] + \textit{symmetric terms}$ and note that $\phi_i^{(I)}(y) = G_{h,y}(Y_i, X_i) K_h(Y_i - y)$. Further, rewrite $\phi_i^{(II)}(y) = \delta^{II}(y)\Omega_i$. We have:

$$
\begin{aligned}
E[|\tilde{\phi}_i^{(I)}(y)|^3] &= E\left[E[|G_{h,y}(Y_i, X_i)|^3 | Y_i] K\left(\frac{Y_i - y}{h}\right)^3\right] \\
&= h \int E[|G_{h,y}(y + uh, X_i)|^3 | y + uh] K^3(u) f(y + uh - X_i'\beta) du \\
&\leq h(1 + \sup_y E|G_{h,y}|^4) \|f\|_\infty \int K(u)^3 du \\
E[|\tilde{\phi}_i^{(II)}(y)|^3] &\leq h^3 E[|\delta^{II}(y)|^3](1 + E\|\Omega_i\|^4)
\end{aligned}
$$

and $\sup_y E|G_{h,y}|^4 = O(1)$ since integration over Y_j and Y_k absorbs the h^{-2} term, also one can derive that $E[|\delta^{II}(y)|^3] = O(1)$ using change of variables and $\int K'(u)du = 0, \int K'(u)udu = -1$. Note that $E\|\Omega_i\|^4 = O(1)$ by the assumptions of Theorem 3. Similar reasoning shows that $E[|\tilde{\phi}_i^{(I)}(y)|^4 \leq hO(1)$ and $E[|\tilde{\phi}_i^{(II)}(y)|^4] \leq h^4 O(1)$.

Furthermore, note that $\|\sup_y \tilde{\phi}_{iy}^{(I)}\|_{P,q} < \infty$ for any q as $\tilde{\phi}_{iy}^{(I)}$ is uniformly bounded and under the assumption that X_i and Ω_i have finite fourth moment we also have $\|\sup_y \tilde{\phi}_{iy}^{(II)}\|_{P,4} < \infty$. Thus, we can apply Lemma 3 with $q = 4, K_n = \log n, \gamma = (\log n)^{-1}, \sigma = h^{1/2}$ and $b = O(1)$ to obtain that there exists a random variable $\tilde{Z}^*$ such that $|\sup_y \sqrt{n} P_n^*(\tilde{\phi}_{iy}^{(I)} + \tilde{\phi}_{iy}^{(II)}) - \tilde{Z}^*| = O_p(n^{-1/6}h^{1/3}\log n + n^{-1/4}h^{1/4}\log^{7/4} n + n^{-1/2}\log^2 n)$, which implies:

$$
\begin{aligned}
|\sup_y \{B(y) + \sqrt{n} P_n^* \phi_i^c(y)\} - Z^*| &= O_p(n^{-1/6}h^{-2/3}\log n + n^{-1/4}h^{-3/4}\log^{7/4} n + n^{-1/2}h^{-1}\log^2 n) \\
&= o_p(1)
\end{aligned}
$$

where the last equality follows from the rate condition in Assumption 1(b) and Z^* follows (conditionally on the sample) the same distribution as $\sup_y\{B(y) + \mathbb{G}_P \phi_i^c(y)\}$, i.e. $P_n^* \mathbb{1}\{Z^* \leq t\} = P\mathbb{1}\{\sup_y\{B(y) + \mathbb{G}_P \phi_i^c(y)\} \leq t\}$ for all t. Similarly, Theorem 2 and Lemma 3 imply also that $|P(S_n^{conc} \leq t) - P(\sup_y\{B(y) + \mathbb{G}_P \phi_i^c(y)\} \leq t)| = o(1)$ for all t. These results imply:

$$
|P(S_n^{conc} \leq c_\alpha^{conc,*}) - P(Z^* \leq c_\alpha^{conc,*})| = |P(S_n^{conc} \leq c_\alpha^{conc,*}) - \alpha)| = o(1)
$$

which concludes the proof.

Let us finish with a note about the result for the global statistic. Showing that $S_n^* - \sqrt{n}\theta$ converges weakly P-almost surely to a normal random variable involves very similar arguments to the ones given above. Then, using this result and Theorem 1, we obtain the final result.

References

Abrevaya, J., Hausman, J. A. & Khan, S. (2010), 'Testing for causal effects in a generalized regression model with endogenous regressors', *Econometrica* **78**(6), 2043–2061.

Abrevaya, J. & Jiang, W. (2005), 'A nonparametric approach to measuring and testing curvature', *Journal of Business & Economic Statistics* **23**(1), 1–19.

Card, D., Chetty, R. & Weber, A. (2007), 'The spike at benefit exhaustion: Leaving the unemployment system or starting a new job?', *American Economic Review* **97**(2), 113–118.

Chen, S. (2002), 'Rank estimation of transformation models', *Econometrica* **70**(4), pp. 1683–1697.

Chernozhukov, V., Chetverikov, D. & Kato, K. (2014), 'Gaussian approximation of suprema of empirical processes', *The Annals of Statistics* **42**(4), 1564 – 1597.

Chernozhukov, V., Chetverikov, D. & Kato, K. (2016), 'Empirical and multiplier bootstraps for suprema of empirical processes of increasing complexity, and related Gaussian couplings', *Stochastic Processes and their Applications* **126**(12), 3632–3651.

Chetverikov, D. (2019), 'Testing regression monotonicity in econometric models', *Econometric Theory* **35**(4), 729–776.

Ekeland, I., Heckman, J. J. & Nesheim, L. (2004), 'Identification and estimation of hedonic models', *Journal of Political Economy* **112**(S1), S60–S109.

Ghosal, S., Sen, A. & van der Vaart, A. W. (2000), 'Testing monotonicity of regression', *The Annals of Statistics* **28**(4), 1054–1082.

Gine, E. & Zinn, J. (1990), 'Bootstrapping general empirical measures', *The Annals of Probability* **18**(2), 851–869.

Gutknecht, D. (2016), 'Testing for monotonicity under endogeneity: An application to the reservation wage function', *Journal of Econometrics* **190**(1), 100–114.

Hainmueller, J., Hiscox, M. J. & Sequeira, S. (2015), 'Consumer demand for fair trade: Evidence from a multistore field experiment', *The Review of Economics and Statistics* **97**(2), 242–256.

Hall, P. & Keilegom, I. V. (2005), 'Testing for monotone increasing hazard rate', *The Annals of Statistics* **33**(3), 1109–1137.

Han, A. K. (1987), 'A non-parametric analysis of transformations', *Journal of Econometrics* **35**(2-3), 191–209.

Ichimura, H. (1993), 'Semiparametric least squares (SLS) and weighted least squares estimation of single index models', *Journal of Econometrics* **58**, 71–120.

Jessoe, K. & Rapson, D. (2014), 'Knowledge is (less) power: Experimental evidence from residential energy use', *American Economic Review* **104**(4), 1417–38.

Karlan, D. S. & Zinman, J. (2008), 'Credit elasticities in less-developed economies: Implications for microfinance', *American Economic Review* **98**(3), 1040–68.

Karlan, D. & Zinman, J. (2018), 'Long-run price elasticities of demand for credit: Evidence from a countrywide field experiment in Mexico'. Working paper.

Kline, B. (2016), 'Identification of the direction of a causal effect by instrumental variables', *Journal of Business & Economic Statistics* **34**(2), 176–184.

Komarova, T. & Hidalgo, J. (2020), 'Testing nonparametric shape restrictions'. Working paper.

Neumeyer, N., Noh, H. & Van Keilegom, I. (2016), 'Heteroscedastic semiparametric transformation models: Estimation and testing for validity', *Statistica Sinica* **26**, 925–954.

Pakes, A. & Pollard, D. (1989), 'Simulation and the asymptotics of optimization estimators', *Econometrica* **57**(5), 1027–57.

Serfling, R. J. (1980), *Approximation theorems of mathematical statistics*, Wiley, Chichester; New York.

Sherman, R. P. (1994), 'Maximal inequalities for degenerate U-processes with applications to optimization estimators', *Ann. Statist.* **22**(1), 439–459.

Subbotin, V. (2007), Asymptotic and bootstrap properties of rank regressions. MPRA Working Paper 9030.

Szydłowski, A. (2020), 'Testing a parametric transformation model versus a nonparametric alternative', *Econometric Theory* **36**(5), 871–906.

Van der Vaart, A. W. & Wellner, J. A. (1996), *Weak Convergence and Empirical Processes: with Applications to Statistics*, Springer-Verlag, New York.